View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Bowling Green State University: ScholarWorks@BGSU

Bowling Green State University

ScholarWorks@BGSU

Chemistry Faculty Publications

Chemistry

7-1998

The 5S rRNA Loop E: Chemical Probing and Phylogenetic Data Versus Crystal Structure

Neocles B. Leontis
Bowling Green State University, leontis@bgsu.edu

Eric Westhof

Follow this and additional works at: https://scholarworks.bgsu.edu/chem_pub

 Part of the Chemistry Commons

Repository Citation

Leontis, Neocles B. and Westhof, Eric, "The 5S rRNA Loop E: Chemical Probing and Phylogenetic Data Versus Crystal Structure" (1998). *Chemistry Faculty Publications*. 12.
https://scholarworks.bgsu.edu/chem_pub/12

This Article is brought to you for free and open access by the Chemistry at ScholarWorks@BGSU. It has been accepted for inclusion in Chemistry Faculty Publications by an authorized administrator of ScholarWorks@BGSU.

RNA (1998), *4*:1134–1153. Cambridge University Press. Printed in the USA.
Copyright © 1998 RNA Society.

The 5S rRNA loop E: Chemical probing and phylogenetic data versus crystal structure

NEOCLES B. LEONTIS[1] and ERIC WESTHOF[2]

[1]Chemistry Department, Bowling Green State University, Bowling Green, Ohio 43403, USA
[2]Institut de Biologie Moléculaire et Cellulaire du CNRS, Modélisations et Simulations des Acides Nucléiques, UPR 9002, 15 rue René Descartes, F-67084 Strasbourg Cedex, France

ABSTRACT

A significant fraction of the bases in a folded, structured RNA molecule participate in noncanonical base pairing interactions, often in the context of internal loops or multi-helix junction loops. The appearance of each new high-resolution RNA structure provides welcome data to guide efforts to understand and predict RNA 3D structure, especially when the RNA in question is a functionally conserved molecule. The recent publication of the crystal structure of the "Loop E" region of bacterial 5S ribosomal RNA is such an event [Correll CC, Freeborn B, Moore PB, Steitz TA, 1997, *Cell 91*:705–712]. In addition to providing more examples of already established noncanonical base pairs, such as purine–purine sheared pairings, *trans*-Hoogsteen UA, and GU wobble pairs, the structure provides the first high-resolution views of two new purine–purine pairings and a new GU pairing. The goal of the present analysis is to expand the capabilities of both chemical probing and phylogenetic analysis to predict with greater accuracy the structures of RNA molecules. First, in light of existing chemical probing data, we investigate what lessons could be learned regarding the interpretation of this widely used method of RNA structure probing. Then we analyze the 3D structure with reference to molecular phylogeny data (assuming conservation of function) to discover what alternative base pairings are geometrically compatible with the structure. The comparisons between previous modeling efforts and crystal structures show that the intricate involvements of ions and water molecules in the maintenance of non-Watson–Crick pairs render the process of correctly identifying the interacting sites in such pairs treacherous, except in cases of *trans*-Hoogsteen A/U or sheared A/G pairs for the adenine N1 site. The phylogenetic analysis identifies A/A, A/C, A/U and C/A, C/C, and C/U pairings isosteric with sheared A/G, as well as A/A and A/C pairings isosteric with both G/U and G/G bifurcated pairings. Thus, each non-Watson–Crick pair could be characterized by a phylogenetic signature of variations between isosteric-like pairings. In addition to the conservative changes, which form a dictionary of pairings isosterically compatible with those observed in the crystal structure, concerted changes involving several base pairs also occur. The latter covariations may indicate transitions between related but distinctive motifs within the loop E of 5S ribosomal RNA.

Keywords: chemical probing; 3D structure; isosteric base pairing; modular motifs; phylogenetic analysis; 5S ribosomal RNA

INTRODUCTION

An RNA motif may be defined as a set of RNA sequences that fold into 3D structures sufficiently close to each other, according to appropriate structural and functional criteria, to be considered essentially identical. The structural criteria generally include the path followed by the sugar–phosphate backbone of the RNA in 3D space, the H-bonding patterns between bases, backbone, and solvent atoms, and the geometry of the stacking, as well as other hydrophobic interactions, especially between the bases. Those local interactions maintain the positioning of chemical groups on the surface of the structure, where they are disposed to interact either with other distant regions of the same molecule or with other molecules (proteins, substrates, other RNAs). A modular view of RNA structure anticipates that certain motifs will occur recurrently to mediate crucial interactions in multiple contexts. A new crystal structure of an RNA molecule having a conserved function and

Reprint requests to: Eric Westhof, Institut de Biologie Moléculaire et Cellulaire du CNRS, Modélisations et Simulations des Acides Nucléiques, UPR 9002, 15 rue René Descartes, F-67084 Strasbourg Cedex, France; e-mail: westhof@ibmc.u-strasbg.fr.

structure allows one to examine sequence variations with an eye toward defining the available sequence space sampled by a particular motif.

5S ribosomal RNA is one of the most intensively studied RNA molecules, ever since its secondary structure was inferred by sequence comparisons more than 20 years ago (Fox & Woese, 1975). Although it has been the subject of numerous biochemical, chemical, and physical studies, its function remains unknown (Noller, 1998). It is found in the large subunit of all ribosomes with the exception of small mitochondrial ones. The 5S rRNA of *Escherichia coli* binds three proteins, L25, L18, and L5, and is an integral component of the 50S subunit. 5S RNA possesses a three-way junction with three helical arms, two of which terminate in hairpin loops. As shown in Figure 1, 5S RNA comprises five regular Watson–Crick helices (numbered successively I–V) and five "loop regions" (designated by the letters A–E). These include the two hairpin loops (C and D), the three-way junction (A), and two "internal loops," loop B between helices II and III and loop E between helices IV and V. The loop E region roughly comprises nt 72–81 and 95–104 in the *E. coli* numbering. In bacterial 5S rRNAs, this is a "symmetrical internal loop" (i.e., equal numbers of residues occur in each strand between the Watson–Crick pairings). Loop E is characterized by a large fraction of noncanonical base pairs that nonetheless are surprisingly resistant to single-strand-specific ribonu-

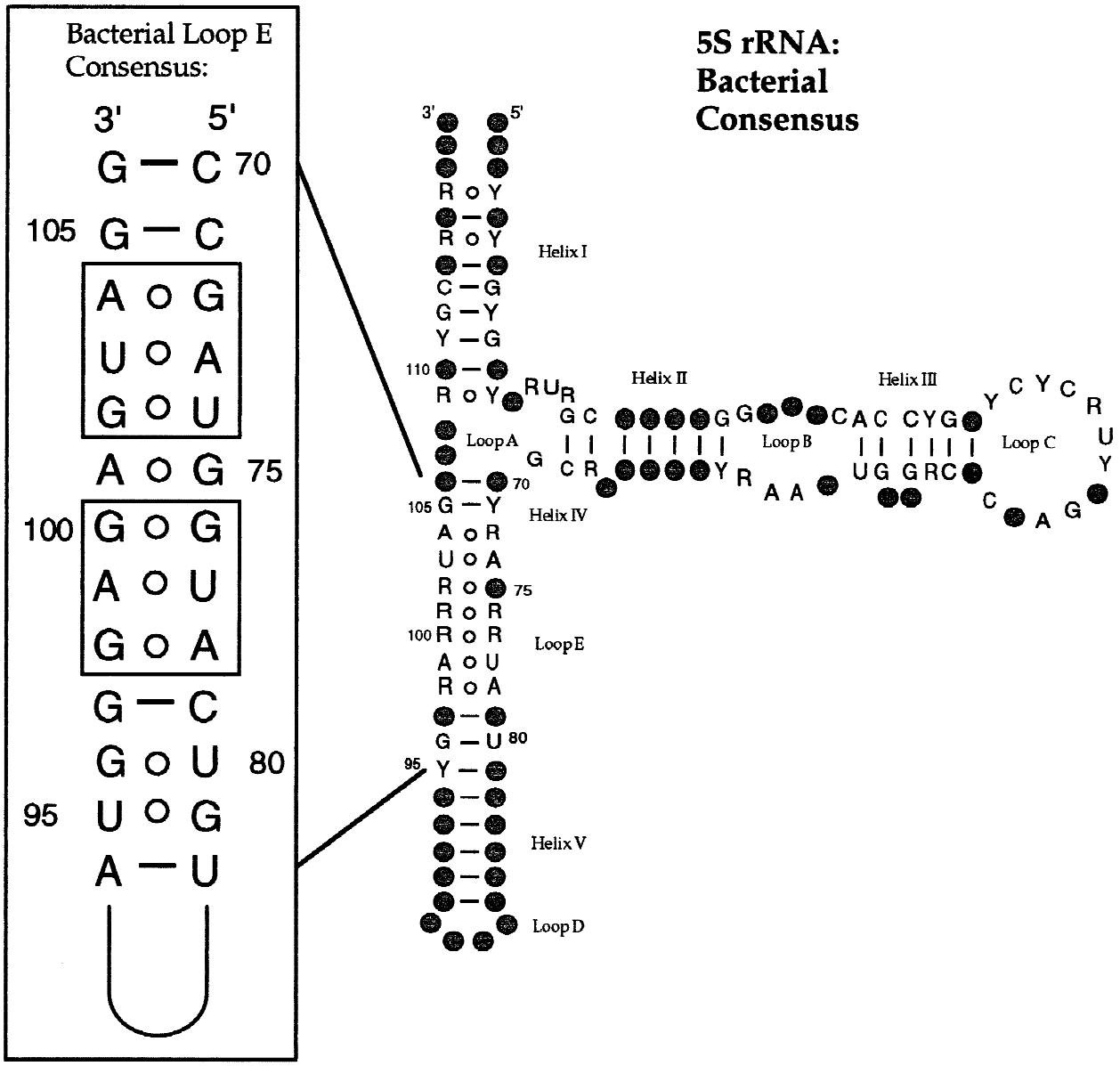

FIGURE 1. Consensus secondary structure of bacterial 5S ribosomal RNA, showing the locations of helices I–V and loops A–E (Szymanski et al., 1998). The consensus sequence of "loop E" (located between helices IV and V), is shown in the inset. Large grey circles: Variable bases; R: A or G; Y: C or U. Small circles indicate noncanonical base pairs.

clease probes (Brunel et al., 1991). In eucarya and one group of the archaea (euryarchaea), loop E forms an asymmetric loop identical in 3D structure to the sarcin–ricin loop of 23S rRNA (Wimberly et al., 1993; Szewczak & Moore, 1995). The eucaryal and bacterial versions of loop E are related but distinct in structure (Romby et al., 1990).

The structure of an RNase-resistant fragment ("Fragment I") comprising nt 1–11, 69–87, and 90–120 of the 5S rRNA of *E. coli* has been solved by X-ray crystallography at 3.0-Å resolution and, in the same study, an RNA duplex containing the minimal 11 base pairs (nt 70–81 and 96–107) needed to form loop E and to bind ribosomal protein L25 has been solved to 1.5-Å resolution (Correll et al., 1997). The appearance of this high-resolution view of the loop E structure provides the opportunity to reassess the extensive chemical and enzymatic probing data as well as the phylogenetic data, to extract new insight into RNA structure and evolution.

The organization of the paper is as follows. First, we will compare the chemical probing data with the crystal structure. This will be done mainly through accessibility calculations of the crystal structure. The accuracy and relevance of those calculations for deducing base pairing geometries will be discussed. Such comparisons will allow us to understand better the limitations and errors of previous modeling efforts. We will then attempt, by sequence comparisons, to rationalize the known loop E sequences, by analyzing the conservative changes that maintain isostericity between base pairs. The phylogenetic analysis reveals, however, that geometrical invariance is not necessarily possible at the single-base pair level. In such instances, concerted changes involving several base pairs occur. Finally, those two sets of comparisons based on chemical modification and phylogenetic data allow us to demonstrate the structural homology between the bacterial and the spinach chloroplast loop E structures and to derive a model in agreement with previously published chemical probing data. In the last section, we identify the loop E motif in two other RNAs.

RESULTS

The consensus bacterial sequence of loop E, defined as those bases found most frequently at each position, is shown in Figure 2 (left panel). However, for comparing complex systems, it is not appropriate to use solely consensus sequences; the identity of motifs is best established by comparing the full range of variations at each position in the sequence. The great number of 5S rRNA sequences (888 total species, 332 bacterial species), which are, additionally, distributed rather uniformly between the various phylogenetic groups (Szymanski et al., 1998), allows such an analysis. The approach assumes that, during evolution, the sequences compatible with a given 3D fold are adequately sampled. Fortuitously, the *E. coli* sequence corresponds closely to the consensus bacterial sequence in loop E (Fig. 2, second panel). The one minor exception is the canonical Watson–Crick pairing C97/G79, which is G97/C79 in the consensus. Loop E in bacteria is flanked on the three-way junction side by two canonical base pairs, G106/C70 and G105/C71 (using the *E. coli* numbering), which comprise helix IV, and by a hairpin loop/stem of variable length on the other side. The crystal structure has revealed that all the bases of loop E are paired in spite of the fact that only one canonical Watson–Crick pairing (between positions 79 and 97) occurs within a stretch of 10 base pairs. Three submotifs exhibiting cross-strand purine–purine stacking were identified in the crystal structure: A104/G72 stacking on U103/A73, A78/G98 stacking on U77/A99, and G96/U80 stacking on U95/G81 (Correll et al., 1997). In Figure 2, boxes are drawn around the two major submotifs of the bacterial loop E motif. As discussed below, sequence comparisons and close examination of the 3D structure show that these two submotifs are essentially identical (submotif 1: G72-U74/G102-A104; submotif 2: G76–A78/G98–G100).

Chemical probing data versus crystal structure

Each non-Watson–Crick base pair comprising the loop E region will be discussed in turn with regard to the H-bonding pattern, the bound water and magnesium ions observed in the crystal structure, and the reactivity toward chemical probes. The probing data from Brunel et al. (1991) is summarized with color coding in Figure 2 (second and fourth panels), with red indicating reactivity under native conditions (in the presence of Mg^{2+}, 20 °C); green, reactivity under semi-denaturing conditions (absence of Mg^{2+}, 20 °C); and blue, reactivity only under denaturing conditions (absence of Mg^{2+} and elevated temperature). The accessibilities of reactive positions as calculated from the crystal structure, with and without bound water and Mg^{2+} ions, are presented in Table 1. The color coding is consistent with the probing data shown in Figure 2.

A104/G72

This sheared purine–purine pair in the crystal structure involves H-bonding in the shallow groove between AN7 and GN2 and between AN6 and GN3. As expected, the N1 of A104 is very reactive, whereas N7 is unreactive. Neither the N1 nor the N7 of G72 are reactive, except under denaturing conditions. This pair, displayed with the water molecules observed in the crystal structure (W), is shown in Figure 3 (upper left). Note the water molecules H-bonded directly to N1, N2, and N7 of G72.

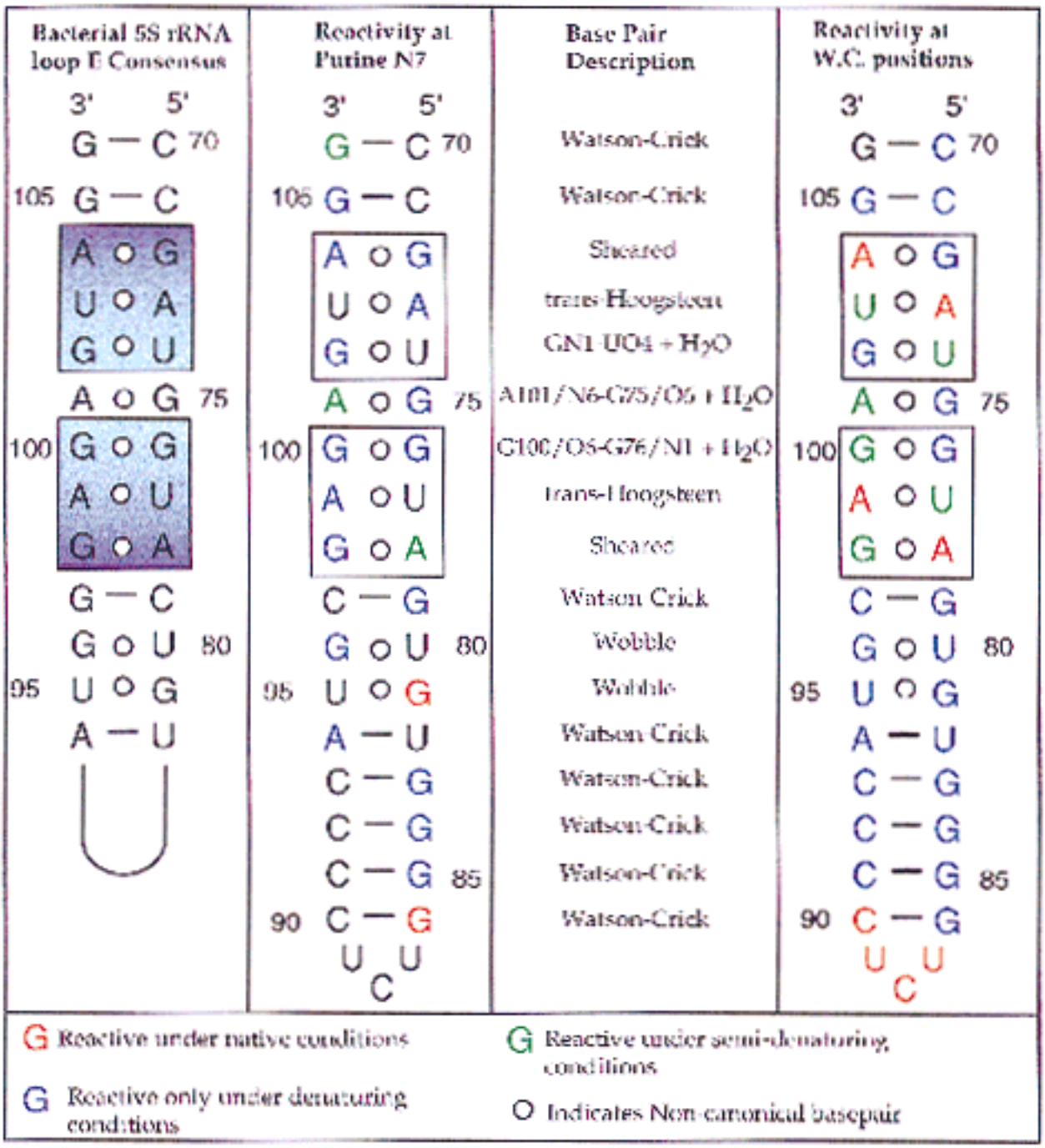

FIGURE 2. First panel: Consensus bacterial 5S rRNA loop E sequence. Relative orientations of the two geometrically very similar submotifs of loop E are indicated by the color gradient. Second panel: Reactivities at purine N7 positions of *E. coli* 5S rRNA (Brunel et al., 1991). Third panel: Description of base pairings (Correll et al., 1997). Fourth panel: Reactivities at Watson–Crick N1 or N3 positions of *E. coli* 5S rRNA (Brunel et al., 1991). Reactivities are color coded as: red, reactive under native conditions; green, reactive under semi-denaturing conditions; blue, reactive under denaturing conditions. Native conditions: Neutral pH, 20 °C, presence of Mg^{2+} ions. Semi-denaturing conditions: Neutral pH, 20 °C, absence of Mg^{2+} ions. Denaturing conditions: 90 °C, absence of Mg^{2+} ions. The two larger boxes delineate the two symmetrically disposed, geometrically identical submotifs of loop E (submotif 1: G72–U74/G102–A104; submotif 2: G76-A78/G98-G100). The color gradient indicates the opposite relative orientation of the submotifs.

The water molecule H-bonded to G72-N7 also contacts an anionic oxygen of the C71 phosphate group, whereas the water H-bonded to G72-N1 and N2 contacts the A104 phosphate group.

U103/A73

This pairing is *trans*-Hoogsteen with H-bonding between AN6 and UO2 and between AN7 and UN3

TABLE 1. Accessibilities of reactive atoms on the bases calculated using the Access program for the loop E of *E. coli* 5S rRNA.[a]

N7		N1/N3		5S rRNA		N1/N3		N7	
1.4 Å/2.8Å		1.4 Å/2.8Å		Loop E		1.4 Å/2.8Å		1.4 Å/2.8Å	
Ions	No ions	Ions	No ions			Ions	No ions	Ions	No ions
0.0	0.0	0.0	6.7	A104	G72	0.9	6.5	0.0	11.1
0.0	0.0	0.0	3.9			0.0	8.0	0.0	5.2
		0.0	0.0	U103	A73	1.9	12.1	0.0	0.0
		0.0	0.0			0.0	6.8	0.0	0.0
0.0	11.1	0.0	0.1	G102	U74	0.0	2.6		
0.0	0.0	0.0	0.0			0.0	0.0		
0.0	2.8	0.0	1.3	A101	G75	0.0	1.3	0.0	7.8
0.0	0.0	0.0	0.0			0.0	0.0	0.0	0.0
0.0	1.2	0.0	5.0	G100	G76	0.0	0.0	0.0	9.8
0.0	0.0	0.0	0.0			0.0	0.0	0.0	0.0
0.0	0.0	7.9	10.2	A99	U77	0.0	0.0		
0.0	0.0	4.0	5.3			0.0	0.0		
0.1	4.6	0.1	7.9	G98	A78	5.0	5.9	0.0	0.0
0.0	3.0	0.0	2.9			0.3	1.8	0.0	0.0
		0.0	0.0	C97	G79	0.0	0.0	0.0	4.8
		0.0	0.0			0.0	0.0	0.0	0.0
0.2	6.6	0.0	0.0	G96	U80	0.0	0.0		
0.0	0.0	0.0	0.0			0.0	0.0		
		0.1	0.1	U95	G81	0.0	0.0	0.3	8.6
		0.0	0.0			0.0	0.0	0.0	2.1

[a]For each residue, accessibilities were calculated using a 1.4-Å radius probe (first line) and a 2.8-Å radius probe (second line). For residues G72–U80 and C97–A104, accessibilities were calculated with coordinates from the 12-mer oligonucleotide (URL064.pdb). For residue G96, accessibilities were calculated with coordinates from the Fragment 1 crystal structure (URL069.pdb). Residues are color coded according to reactivities reported by Brunel et al. (1991). Blue, reactive only under denaturing conditions; green, reactive under semi-denaturing conditions (absence of Mg^{2+} ions); red, reactive under native conditions.

(Fig. 4). The N1 of A73 is reactive under native conditions, whereas N7 is unreactive, as expected from the H-bonding pattern. N3 of U103 is reactive, but only under semi-denaturing conditions.

G102/U74

This pairing is not of the typical wobble variety. The O4 of U74 is H-bonded to both N1 and N2 of G102 in a bifurcated fashion (see Fig. 5). A water molecule bridges G102-N2 and U74–N3. The N3 of U74 is reactive under semi-denaturing conditions, whereas the N1 and N7 of G102 only react under denaturing conditions. The crystal structure reveals water molecules bound to G102-N7 and G102-O6. Two of these water molecules are coordinated by a magnesium ion (belonging to the "tandem pair" described below) located in the the deep groove between this base pair and A101/G75.

A101/G75

This and the next base pair exhibit new geometries (A101/G75, Fig. 6; G100/G76, Fig. 5) with A101-N6 H-bonded to G75-O6 and a water molecule bridging A101-N1 with G75-N1. The N1 and N7 of A101 are unreactive under native conditions, but become reactive in semi-denaturing conditions, whereas N1 and N7 of G75 remain unreactive. Two magnesium ions, bridged by three water molecules, form a tandem pair in the deep groove. One ion coordinates directly to the phosphate of G100, whereas the other coordinates to the phosphate of A101. Water molecules directly coordinated by these Mg^{2+} ions H-bond to the 6 and 7 positions of each of the stacked purines G100, A101, and, as mentioned above, to G102. A third Mg^{2+} ion coordinates the four water molecules directly contacting the O6 and N7 positions of G75 and G76.

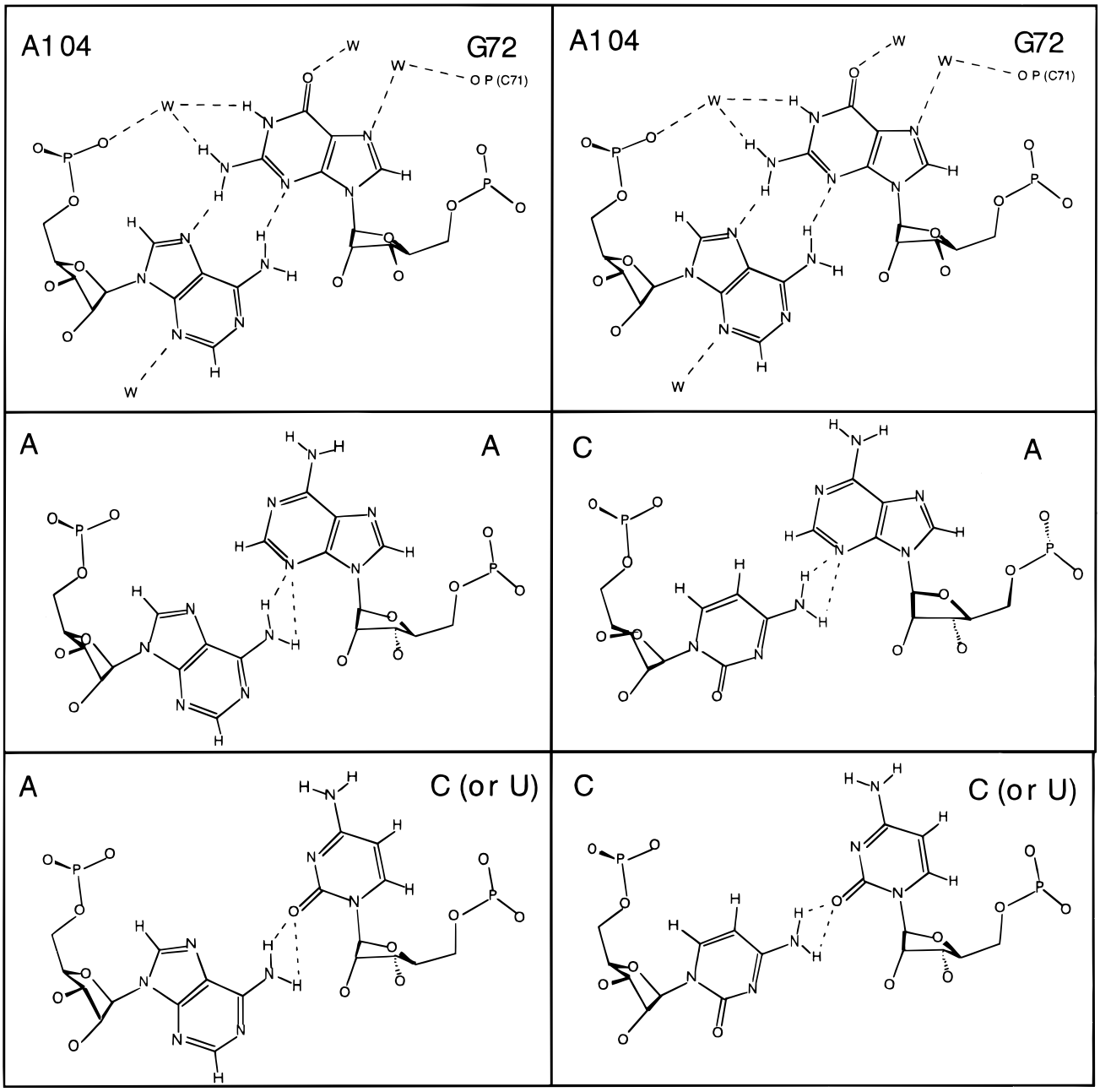

FIGURE 3. Pairing 104/72. Upper panels: A104/G72 pairing as observed in the crystal structure of the 12-mer oligonucleotide corresponding to loop E of *E. coli* 5S rRNA (URL064.PDB). Water molecules are indicated by "W." Dashed lines indicate H-bonds. Left panels: Isosteric pairs retaining A104 observed as conservative substitutions in the 5S rRNA bacterial database (Szymanski et al., 1998). Right panels: Isosteric pairs substituting C104 observed as conservative substitutions in the 5S rRNA bacterial database. In Figures 3, 4, 5, and 6, the drawings of the isosteric pairs are only indicative because, purposefully, no energy minimizations were performed. In most cases, it is apparent that small variations would improve the H-bonding and van der Waals geometries.

G100/G76

This purine–purine pairing involves a bifurcated H-bond between the O6 of G100 and the N1 and N2 of G76. A water molecule bridges G76-N2 and G100-N1. A water molecule also bridges G100-N7 and G76-O6. Only the N1 of G100 is reactive, and only under semi-denaturing conditions. The two water molecules bonded to G76-N7 and O6 in the plane of the G100/G76 base pair are directly coordinated by a magnesium ion, as already mentioned.

The similarity of this pairing to the consensus G102/U74 pairing is noteworthy (see Fig. 5). Both involve a bifurcated H-bond between N1 and N2 of a guanosine

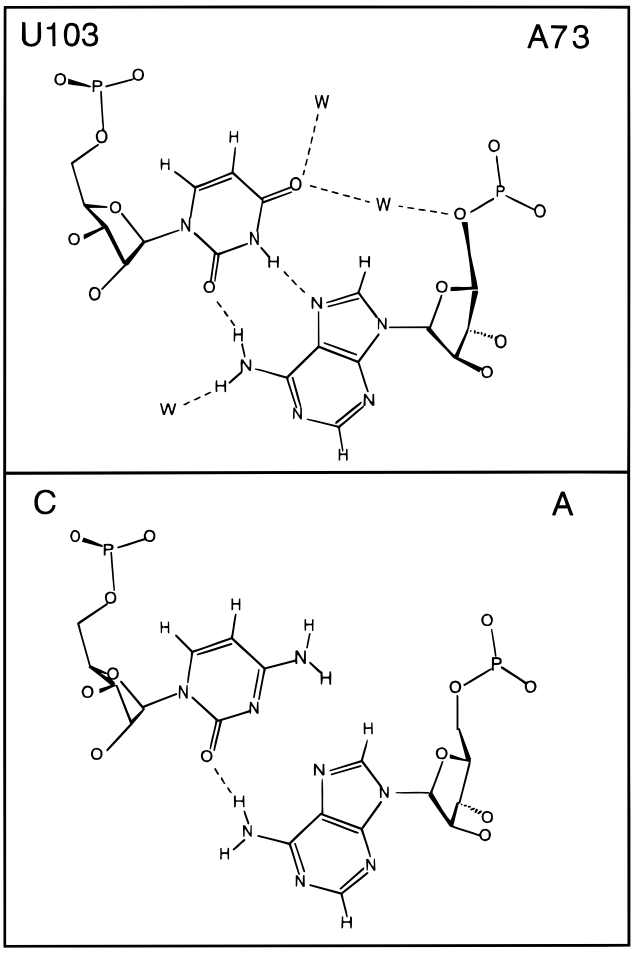

FIGURE 4. Pairing 103/73. Top panel: Highly conserved U103/A73 pairing as observed in the crystal structure of the 12-mer oligonucleotide (see Fig. 3 legend). Lower panel: Isosteric pair C103/A73 observed as conservative substitution in the 5S rRNA bacterial database (Szymanski et al., 1998).

base and a carbonyl belonging either to a uridine or to another guanosine. In each case, a water molecule bridges from N2 of the guanosine to the imino nitrogen of the uridine or guanosine partner, completing the base pairing. Similar H-bonding patterns are present in the P5abc of the Group I intron (U168/G188) (Cate et al., 1996) and in the tRNA structure (Saenger, 1984), where they involve tertiary pairs (e.g., G45 and U25 or G18 and Ψ55). The observation that the phosphodiester backbones are locally parallel in the G18/Ψ55 base pair suggests that "bifurcated" G/U pairs could occur with a locally parallel orientation of the strands (with O2 of U interacting with N1 and N2 of G) and that "bifurcated" G/Ψ could also occur with a locally antiparallel orientation. The O4 of Ψ55 is positioned to form a bifurcated H-bond with N1 and N2 of G18 while exposing the N3 of Ψ55. A phosphate oxygen (A58), positioned at a site equivalent to the water molecules in the two loop E pairings H-bonds to ΨN3.

A99/U77

This *trans*-Hoogsteen pairing displays exactly the same geometry and reactivity as observed for U103/A73 (Fig. 4).

G98/A78

This sheared purine–purine base pair is equivalent in geometry to A104/G72 (Fig. 3). Only AN1 is reactive under native conditions. AN7 and GN1 are reactive under semi-denaturing conditions. A magnesium ion is observed directly coordinated to O6 of G72. The water molecule contacting G72-N7 is coordinated to this same magnesium.

C97/G79

This is the sole Watson–Crick pair within loop E. Watson–Crick and purine N7 positions are only reactive under denaturing conditions.

G96/U80

This pairing exhibits a typical wobble geometry in which GN1 is H-bonded with UO2 and GO6 is H-bonded with UN3. G96 is stacked on G81 in a cross-strand manner, similar to A104 stacking on A73 and A99 stacking on A78. Neither the Watson–Crick positions nor GN7 are reactive, except under denaturing conditions.

U95/G81

In the crystal structure, this is a wobble base pair with identical H-bonding as G96/U80. However, as a result of the cross-strand stacking of G96 and G81, the N7 of G81 is exposed and reactive under native conditions. The Watson–Crick faces of both G and U are, however, unreactive.

Accessibility calculations

The accessibilities of reactive positions in the crystal structures of the helix IV/loop E region of 5S rRNA (URL064.pdb, URL065.pdb, and URL066.pdb) were calculated using the ACCESS program (Richmond, 1984). The calculations were performed with probe radii of 1.4 and 2.8 Å, with and without Mg^{2+} ions and bound water present (Table 1). At present, it remains theoretically unfounded to attempt to predict water-binding sites around a modeled structure. Thus, in order to be useful for assessing the accuracy of a modeled structure, the accessibility calculations performed in the absence of water molecules and ions should correlate with modification data. As discussed previously (Westhof et al., 1989), it is appreciated that accessibility alone is an insufficient theoretical measure of chemical reactivity.

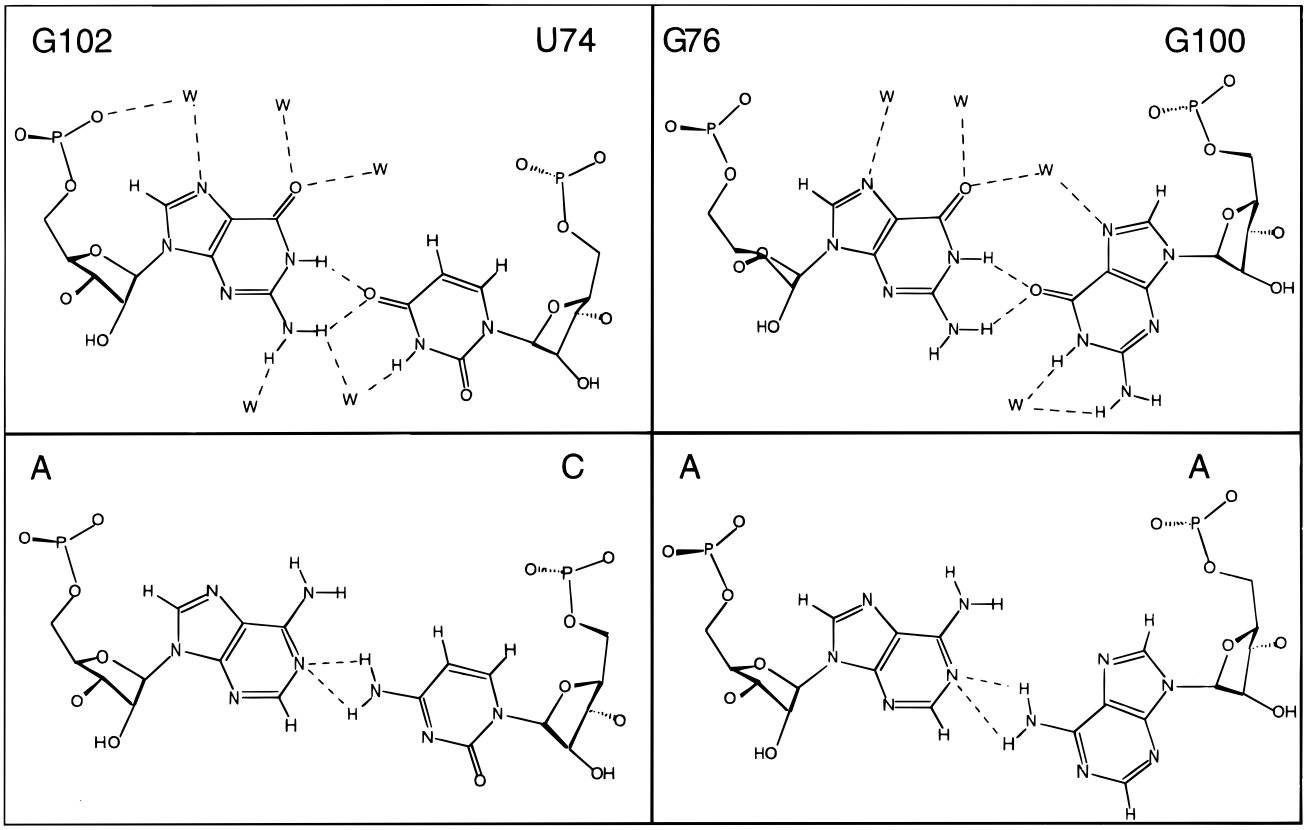

FIGURE 5. Isosteric bifurcated pairings 102/74 and 76/100. Top panels: Bifurcated G102/U74 (left) and G76/G100 (right) pairings as observed in the crystal structure of the 12-mer oligonucleotide (see Fig. 3 legend). Note the equivalent role played by the carbonyl groups of U74 and G100 in forming bifurcated H-bonds to their partners, G102 and G76, respectively. Lower panels: Isosteric substitutions observed for each pairing in the 5S rRNA bacterial database (Szymanski et al., 1998).

First of all, while lack of accessibility precludes reactivity, adequate accessibility does not necessarily lead to reactivity. Further, because the calculations are made on a static molecule, their value for interpreting reactivities under semi-denaturing conditions (which reflect increased molecular mobilities) is limited.

Watson–Crick positions

The N1/N3 positions of all canonically paired bases, unreactive except under denaturing conditions in chemical probing experiments, are found to be inaccessible (exposed surface area equal to zero), with or without ions and water molecules, in the crystal structures. The only bases that are reactive under native conditions at their Watson–Crick H-bonding positions are the adenosines A73, A78, A99, and A104. Two of these (A78 and A104) are involved in sheared A/G pairings and the other two (A73 and A99) in A/U *trans*-Hoogsteen pairings, both pairing schemes having the reacting atom N1 not involved. Significant accessibilities (>5 Å^2) are calculated for N1 of all of these (in the absence of ions and water molecules) using the 1.4-Å radius probe.

The following bases show reactivity at N1 or N3 under semi-denaturing conditions: U74, U77, G98, G100, A101, and U103. The N1 of G100, the N3 of U74, and the N1 of A101 (Figs. 5, 6), all show some weak (<5 Å^2) accessibility to the 1.4-Å probe when water and ions are removed. However, U77 and U103, both of which are involved in U/A *trans*-Hoogsteen pairings, are calculated as inaccessible in the absence of ions and water molecules. This suggests that removal of magnesium ions from the solution destabilizes the loop E structure so as to expose the Watson–Crick positions of these bases. The N1 atoms of G75, G76, and G102 remain unreactive, even under semi-denaturing conditions, and are found to be inaccessible. In short, the agreement between the modification data of Watson–Crick positions, the accessibility calculations, and the crystal structure is rather good. “Red” residues (N1’s of A’s in *trans*-Hoogsteen A/U and sheared A/G pairs) present high accessibilities, “green” residues are generally intermediate, and “blue” sites present close to zero accessibilities. Notice also that with reactivities at Watson–Crick positions, the size of the probe does not appear to have an effect.

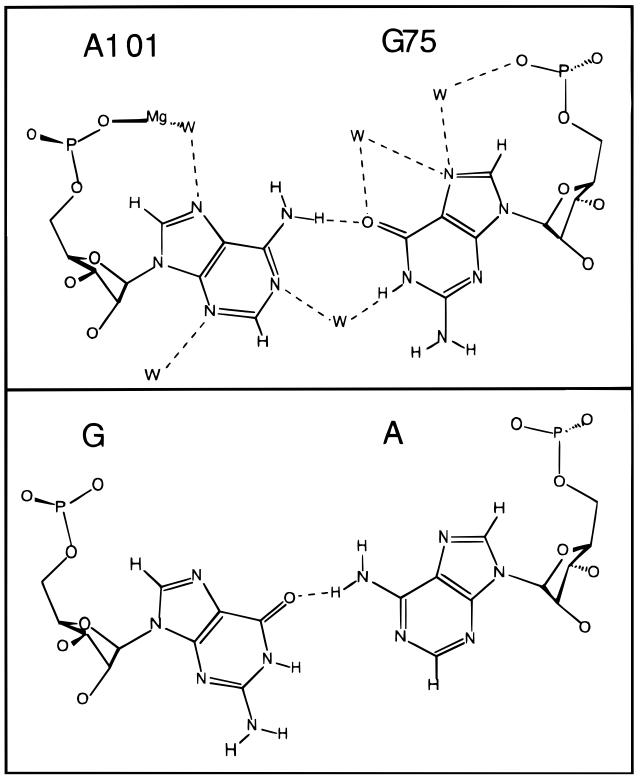

FIGURE 6. Isosteric pairings A101/G75 and G101/A75. Top panel: A101/G75 pairing as observed in the crystal structure of the 12-mer oligonucleotide (see Fig. 3 legend). Lower panels: Isosteric substitution G101/A75, observed to covary with A101/G75 in the 5S rRNA bacterial database (Szymanski et al., 1998).

The N1's of G72 and G98 present accessibilities too high for their lack of reactivity. In those two instances (both involved in sheared A/G pairs), the calculations have to include water and ions to yield the low accessibilities expected from the lack of chemical modification. It is worth noting that the water molecule bound to G98-N1 and N2 is also directly contacted by an anionic oxygen of the U77 phosphate group. The equivalent water molecule bound to G72-N1/N2 is contacted by both the U103 and the A104 phosphate groups. These two water molecules have *B*-factors that are significantly below the average over all water molecules in the structure (28.1 Å^2 and 28.8 Å^2 compared to an average *B*-factor of 34.7 $\pm$ 5.8 Å^2 for all water molecules, an average *B*-factor of 24.8 $\pm$ 4.2 Å^2 for the five water molecules participating directly in the bifurcated G/G, G/U, and open A/G base pairs, and finally an average *B*-factor of 21.7 $\pm$ 2.7 Å^2 for all RNA atoms in the structure).

Purine N7 sites

For the N7 positions, on the other hand, the size of the probe has a significant effect on the accessibilities. Attention should be paid to the trends in both values. The only purine N7 position found to be reactive under native conditions is G81, which belongs to the GU tandem pair submotif of loop E and which predicts accessibilities for both probe sizes. The N7 positions unreactive in native or reactive in semi-denaturing conditions present zero or close to zero accessibilities with the larger probe size, except for two residues, G72 and G98, both involved in sheared A/G pairs. Direct comparison of the *B*-factors of the water molecules bound to the N7's of G72 and G98 is not warranted because a Mg^{2+} ion is directly coordinated to one (G98), but not to the other. In fact, Mg^{2+} ions are directly coordinated to the water molecules contacting the N7's of 7 of the 11 G's in the structure. Therefore, it is understandable that for the two residues G72 and G98, the ion and water shell have to be included to obtain full inaccessibility expected from their low experimental reactivities at N7.

In short, these comparisons reveal two trustworthy signatures: one for the *trans*-Hoogsteen A/U and one for the sheared A/G pairs. A *trans*-A/U Hoogsteen pair presents a reactive A-N1 with a protected A-N7 and the U-N3 reactive in semi-denaturing conditions. A sheared A/G pair presents a reactive A-N1, with an A-N7 either protected in native or reactive in semi-denaturing conditions. In the sheared A/G pair, both the N1 and N7 of the G are protected in the native state. A similar pattern of chemical reactivity was observed for the tandem sheared A/G pairs in the conserved quartet of the SECIS element (Walczak et al., 1996).

Overall, the agreement between the X-ray structure and the chemical probing reactivities is adequate. The main exception is the G residue (at both N1 and N7 positions) of sheared A/G pairs. We suggest that the water and ion shell surrounding the sheared A/G pairs hinders the reactivity of guanine residues (see Fig. 3). NMR of DNA fragments (Dennisov et al., 1997) as well as molecular dynamics simulations of RNAs (Auffinger & Westhof, 1997) have shown that water molecules in some hydration sites reside for hundreds of picoseconds. The diffusing reactive species could well interact with water molecules before the water molecules in direct contact with the G residues exchange with the bulk solvent, especially when the observed "bound water" presents a low *B*-factor (i.e., a low temperature factor and thus a high occupancy and a long residence time). A well-documented example occurs with pseudouridines in tRNAs, whose second imino proton (N1) has a markedly reduced exchange rate with bulk solvent due to a water bridge to its 5′-phosphate (Auffinger et al., 1996; Auffinger & Westhof, 1998; Davis, 1998).

Comparison to previous modeling studies of 5S RNA

Much can be learned by comparing the X-ray structure to previous efforts to model loop E on the basis of the

chemical probing data (Romby et al., 1988; Westhof et al., 1989; Brunel et al., 1991). The probing data indicated loudly and clearly that, in spite of the low proportion of base juxtapositions capable of forming canonical Watson–Crick pairings, loop E is a tight structure in which all bases are paired, at least in the presence of magnesium ions. The modeling studies took this into consideration. The integral part played by bound water molecules in stabilizing noncanonical pairs (Westhof, 1988) was realized, but it was not anticipated that they could complete the H-bonding schemes. Further, although the importance of Mg^{2+} ions was fully appreciated, specific binding modes with their complex water networks were unavailable. This state of affairs, together with a rather dynamic view of Mg^{2+} ion binding, led to other schemes for reproducing the chemical reactivities, including the use of syn-purine base pairs, none of which, however, appear in this structure. Early NMR work on 5S RNA Fragment 1 demonstrated that divalent cations, including Ca^{2+} as well as Mg^{2+}, stabilize the structure of loop E (Leontis et al., 1986). In the absence of divalent cations, all imino proton resonances associated with loop E were found to broaden dramatically or disappear from the spectrum, a result confirmed with a smaller fragment that comprises only the residues of helix V through loop D (Dallas & Moore, 1997).

Sequence analysis

Sequence variations observed in the bacterial 5S rRNA database may be classified as conservative or concerted. We will use the term “conservative change” to refer to the substitution of a single base or base pair by a potentially isosteric pairing without changes in immediately adjacent pairs within a particular sequence of the database. For example, substitution of G by A in a sheared A/G pair is conservative, whereas the reversed base pair G/A is not. Concerted changes, however, refer to instances in which a change of one or more bases in a pair is accompanied by changes in one or both flanking pairs. Analysis of covariations base pair by base pair will be discussed. Phylogenetic data are presented in Table 2.

G106/C70

The G/C pairing appears in 278 of 306 cases in the bacterial database, but sufficient variation is observed to support standard Watson–Crick pairing and to indicate that wobble pairing is not tolerated at this position (Table 2).

G105/C71

G/C pairing is also preferred in 278 of 306 cases (although no correlation exists with the preceding base pair). The sequence variation indicates that wobble pairing is tolerated but Watson–Crick pairing is preferred.

A104/G72

As discussed above, two sheared purine–purine pairings are found in the loop E structure, A104/G72 and A78/G98. Both are expected to exchange with A/A, which can adopt the same geometry, but not with G/A or G/G, which cannot. In fact, A/A is the most commonly observed substitution at both positions 104/72 and 78/98 (Table 2). The following conservative substitutions are observed for the 104/72 sheared pairing: A104/A72, A104/C72, A104/U72, all of which preserve A104; and C104/A72, C104/C72, C104/G72, C104/U72, all of which have C104. Of these, the most frequent pairings in the database are A/G (consensus), A/A, C/C, and C/U, stressing the characteristic H-bond of the sheared base pair between the amino group (N6 of A or N4 of C) and the N3 of a purine or the O2 of a pyrimidine. As shown in Figure 3, all of these pairings can be accommodated isosterically, with the exception of the single C/G, which would require considerable readjustment to accommodate the two amino groups. The A/C opposition is similar to one seen at positions 32 and 38 at the base of the anticodon loop in yeast $tRNA^{Phe}$ (Kim et al., 1973; Sussman et al., 1978). The isosteric structures in Figures 3, 4, 5, and 6 were generated with the FRAGMENT program (Westhof, 1993) using the crystal structure as a template. Interestingly, all alternative pairings, which would necessarily involve U or G at position 104 are not feasible and are not observed. The database thus appears to display all geometrically possible variants for this pairing. In two cases (*Paracoccus denitrificans* and *Rhodobacter sphaeroides*), A104 is deleted while G72 is simultaneously replaced by an A, but the rest of the consensus loop E sequence is maintained. The single U/G pairing at this position belongs to a concerted change in the 5S of *Lactococcus lactis cremoris* (a Gram-positive clostridiobacterium) and will be discussed below. It should not be considered a variant of the motif. The third sequence in which A104 is deleted also involves a concerted change.

U103/A73

This pairing is *trans*-Hoogsteen with H-bonding between AN6 and UO2 and between AN7 and UN3 (Fig. 4). The U/A is almost universally conserved as expected for this special type of pairing. Three cases of C103/A73 pairings are, however, observed. In two cases, those of *Leptonema illini* in the Spirochaetales group and *Bacillus methanolicus* (strain S2A1) in the Clostridiobacteria, this is a conservative change (the sequences otherwise correspond exactly to the con-

TABLE 2. Sequence variations observed at the noncanonically paired positions of loop E of bacterial 5S rRNA.

G106/C70	A	C	G	U	Totals:
A		1 (10.1)	6 (0.54)	10 (0.45)	17
C		0 (14.)	15 (0.00)	1 (0.68)	16
G		279 (254.)	0 (13.68)	1 (10.54)	279
Totals:	0	279	15	12	306

G105/C71	A	C	G	U	Totals:
A	0 (0.04)	0 (0.84)		11 (0.07)	11
G	0 (0.90)	278 (265.19)		15 (25.65)	293
U	1 (0.01)	0 (1.02)		1 (2.18)	2
Totals:	1	278	0	27	306

A104/G72	A	C	G	U	Totals:
A	40 (42.55)	2 (4.55)	224 (210.73)	1 (8.17)	53
C	1 (3.73)	14 (1.23)	1 (6.20)	8 (0.71)	16
G					0
U	0 (0.13)	0 (0.03)	1 (0.76)	0 (0.03)	245
Deletion	2 (0.47)	0 (0.10)	1 (2.28)	0 (0.00)	5
Totals:	48	16	232	9	319

U103/A73	A	C	G	U	Totals:
A	0 (1.30)		1 (0.00)		1
C	3 (2.46)		0 (0.01)		3
G	1 (1.30)		0 (1.30)		1
U	301 (300.01)		0 (0.99)		302
Totals:	305	0	1	0	306

G102/U74	A	C	G	U	Totals:
A	17 (1.50)	7 (0.54)	0 (0.25)	1 (22.55)	25
C	3 (0.19)	0 (0.06)	1 (0.03)	3 (2.71)	8
G	1 (17.20)	1 (7.24)	2 (2.73)	273 (249.84)	277
U	1 (0.36)	0 (0.33)	0 (0.01)	0 (0.00)	1
Totals:	19	8	3	278	306

A101/G75	A	C	G	U	Totals:
A	5 (17.34)	0 (0.94)	230 (237.29)	5 (2.83)	239
C	0 (0.05)	1 (0.00)	0 (0.52)	0 (0.01)	1
G	13 (0.63)	2 (0.05)	0 (13.47)	0 (0.16)	15
U	1 (0.01)	3 (0.00)	0 (0.52)	0 (0.01)	1
Totals:	19	1	282	3	306

S100/G78	A	C	G	U	Totals:
A	93 (95.80)	1 (0.31)	2 (85.88)		96
C	1 (0.31)	0 (0.00)	0 (0.69)		1
G	1 (54.80)	0 (0.69)	296 (43.46)		298
Totals:	95	1	210	0	306

A99/U77	U	C	G	U	Totals:
A	306 (306.00)				306
C					0
U					0
Totals:	306	0	0	0	306

C96/A79	A	C	G	U	Totals
A	32 (32.79)		0 (0.99)	1 (0.11)	33
C	4 (3.06)		0 (0.03)	0 (0.01)	4
G	266 (269.87)		1 (0.67)	0 (0.87)	285
U	3 (2.67)		0 (0.02)	0 (0.01)	3
Deletion	1 (0.86)		0 (0.01)	0 (0.00)	1
Totals	294	0	1	1	306

C97/G79	A	C	G	U	Totals:
A	0 (0.33)	0 (1.57)	0 (1.30)	3 (0.04)	3
C	0 (0.47)	0 (26.76)	51 (24.15)	0 (0.63)	51
G	0 (1.67)	171 (83.70)	0 (77.63)	0 (2.13)	171
U	3 (0.54)	0 (52.98)	97 (45.85)	1 (1.24)	101
Totals:	3	171	148	4	326

S96/U80	A	C	G	U	Totals
A					0
C			11 (11.00)		11
G			1 (1.00)		1
U			288 (286.00)		288
Totals:	0	0	298	0	298

[a]Below the observed frequencies in the 5S rRNA database (Szymanski et al., 1998), the calculated, statistically expected frequencies are shown in parentheses (Chiu & Kolodziedjczak, 1991). Red characters are used to indicate pairings found to be isosteric (or nearly so).

sensus). In the third case (*Flectobacillus major*, in the Cytophagales group of bacteria), the C/A pairing is coupled with deletion of A104, but no other change. The G/A pairing, observed in the *Thermomicrobium roseum* sequence, and the A/G pairing, observed in *L. lactis cremoris*, are both due to concerted changes involving neighboring base pairs, as discussed below. The C/A pairing is shown in Figure 4 (lower panel). It is isosteric with the U/A that it replaces, but lacks one H-bond (should CN3 remain unprotonated), which could explain its low incidence. However, a slight lateral movement of the cytosine residue would lead to the formation of two H-bonds (CN3 to AN6 and CN4 to AN7).

G102/U74

As discussed above, this pairing is not of the typical wobble variety, but closely resembles the G100/G76 pairing, because it also involves a bifurcated H-bond from a carbonyl group (UO4) to the Watson–Crick positions of the guanosine (Fig. 4). The following conservative substitutions are observed: A/A, A/C, A/U, G/A, G/C, and G/G, of which A/A and A/C are statistically favored (Table 2). As shown in Figure 5, A102/A74 can be accommodated by A74 H-bonding via its N6 amino group to the N1 of the adenosine that replaces G102. A slight reorientation could also lead to the formation of an additional A74(N7) to A102(N6) H-bond. A102/C74 is equally easily accommodated using AN1 as the acceptor and CN4 as the H-bond donor in place of the GN1-UO4 H-bond. A water molecule could potentially bridge from C74(N3) or A74(N1) to the polarized base proton H2 of A102 in these pairings.

Two cases of G/G are observed in the database (Table 2). Given the geometrical similarity discussed above between the G100/G76 and the G102/U74 pairing, it is reasonable to propose that the G102/G74 substitution adopts the G100/G76 geometry, with G100 playing the role of U74. In fact, the G100/G76 pair may be superimposed almost exactly upon G102/U74 in 3D space: the sugar–phosphate backbones superimpose precisely, whereas substitution of G74 merely pushes the guanosine base 102 approximately 1.3 Å into the shallow groove side. Thus, submotifs 1 and 2 of loop E, which differ only in the third pairing (G102/U74 in submotif 1 and G76/G100 in submotif 2), are seen to be essentially identical geometrically, as shown in stereo in Figure 7. Both of the G102/G74 substitutions found in the database are coupled with substitutions at the 101/75 pairing position. In one case (*Bacillus acidocaldarius*), G102/G74 is coupled with A101/A75, whereas in the second (*Lactobacillus plantarum*), the G101/A75 pairing (isosteric with the consensus A101/G75 pairing) occurs. The changes at the 101/75 pairing observed in these two cases may reflect changes in hydration and magnesium ion binding.

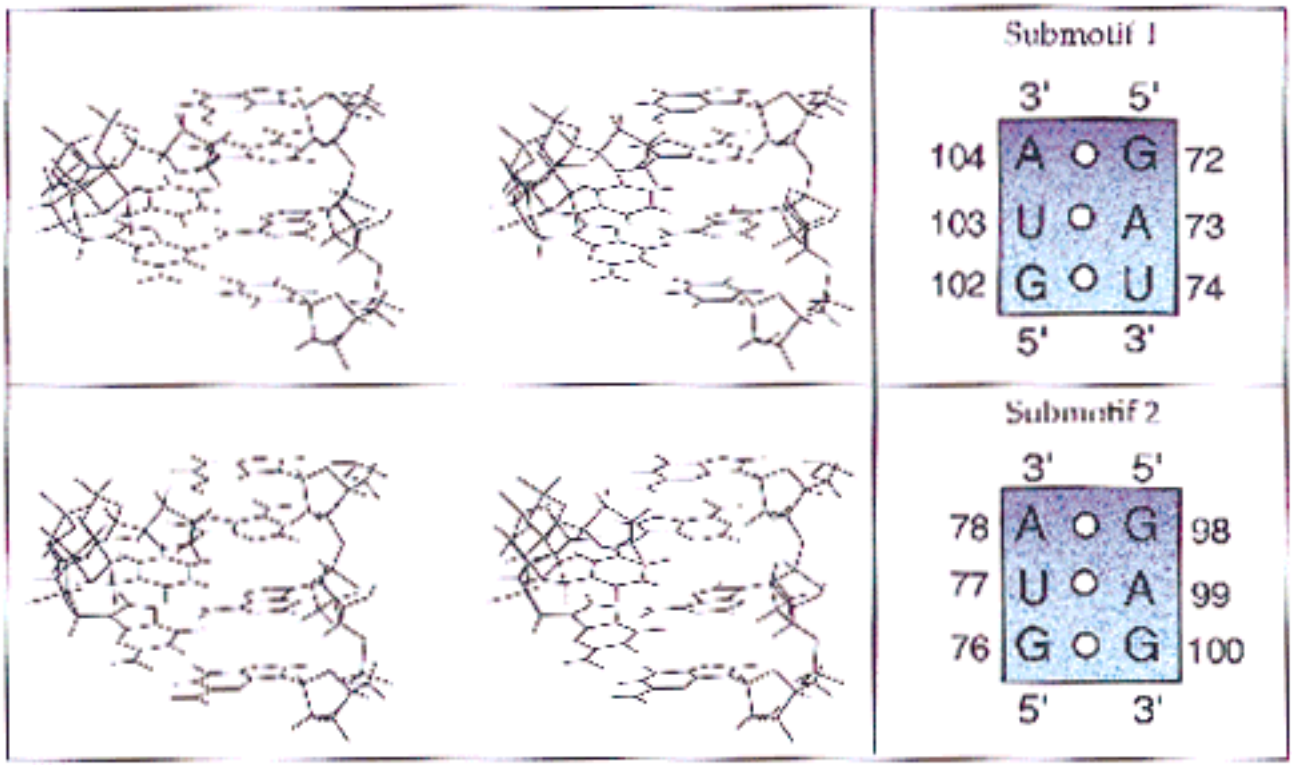

FIGURE 7. Stereo views of submotifs 1 and 2 of loop E (URL064.PDB), oriented to show their geometrical similarity.

The A102/U74 substitution (observed in *Lactobacillus brevis*) is also coupled with a switch in the next pair (to G101/A75). The A102/U74 pairing is geometrically possible (as an isosteric replacement for G102/U74), but less stable than G/U, unless AN1 is protonated. The switch of the 101/75 base pair may again indicate a change in hydration, because magnesium binding is centered around the 101/75 pair in the crystal structures. The G102/A74 (*Sporolactobacillus inulinus*) and G102/C74 (*Verrucomicrobium spinosum*) substitutions are both conservative. However, they cannot exist as isosteric pairings due to the clash of the A or C amino groups with GN1. Large adjustment of geometry must occur to accommodate these pairings.

The two C/U pairings observed (*Anaerorhabdus furcosus* and *Chromatium minutissimum*) both involve concerted changes. The isolated C102/G74 (*T. roseum*) and U102/A74 (*L. lactis cremoris*) variations also involve concerted changes, as mentioned above. Interestingly, there are no instances of U/G or C/A, which would require a different geometry.

A101/G75

This pairing varies almost exclusively with G/A, which is isosteric to it, as may be expected from the symmetry of the pairing geometry (Fig. 6). Although three instances of A/A pairing are also observed, there is a strong statistical bias against homopurine pairings (no cases of G/G pairing are observed). A shift in geometry would be required to prevent clash of the amino groups of the two adenosines, perhaps allowing N6 of one base to pair with N1 of the other, so this is not an isosteric pairing. One case (*B. acidocaldarius*) is coupled to a G102/G74 pairing, which, as discussed above, should be considered a conservative change. The other two involve concerted changes (*Lactococcus* and *Ureaplasma urealyticum*). The absence of G/G substitutions for the A101/A75 pairing is consistent with the geometry. One conservative A101/U75 pairing (*Enterococcus faecalis*) occurs in the database, as well as an instance in which A101/U75 is coupled with A102/A74 (*Mycoplasma pneumoniae*), and another in which it is coupled with C102/U74 (*A. furcosus*). Another case involves larger concerted changes and is discussed below (*U. urealyticum*). An instance of U101/A75 coupled with A102/C74 also occurs (*Bacteroides capillosus*). All the pyrimidine–purine pairings fit, but require water molecules to bridge the H-bonding receptors and acceptors at the Watson–Crick positions of the A and U bases. The C/C pairing belongs to *Chromatium* (concerted change). In summary, the 101/75 pairing gives a clear signature, almost exclusively A/G or G/A. The isosteric G101/A75 pairing may be generated by simply rotating the A101/G75 base pair around the (pseudo-symmetric) axis passing between the bases, perpendicular to the axis of the double helix. The sugar–phosphate backbones of the original and the rotated pair superimpose exactly.

G100/G76

The G/G pair covaries almost exclusively with A/A in the database, with a strong statistical bias against G/A pairs (giving a signature opposite to the preceding pair). As shown in Figure 5 and discussed above, an isosteric pair can be accommodated for this pairing as for the bifurcated G102/U74 pairing by H-bonding of A100-N6 to A76-N1. Note that both the G102/U74 and the G100/G76 pairings covary with A/A. A C100/A76 pairing and a A100/C76 pairing are also observed in the database. The A100/C76 belongs to a concerted change (*Chromatium*), whereas C100/A76 is conservative (*Thiovulum sp.*). C100/A76 is isosteric with G100/G76, whereas A100/C76 is not, just as only A102/C74 is isosteric with the G102/U74 pair, whereas C102/A74 is not. Thus, again, one sees a clear correlation between conservative substitutions and isosteric pairings. Two cases of conservative A100/G76 substitutions occur (*Mycobacterium flavescens* and *Rhodoporphyra umbilicalis*, a plastid). The A100/G76 pairing requires geometrical adaptation, because the amino group of A100 would otherwise be directed toward the imino of G76, as noted above for an equivalent G102/A74 substitution for the G102/U74 pairing. For example, the amino of A100 could pair with the carbonyl of G76, as in the preceding base pair (A101/G75). A single case of a conservative G100/A76 substitution is also found in the database (*Proteus vulgaris*). This sequence was obtained in 1975 using RNA fingerprinting techniques (Fischel & Ebel, 1975). The authors report that portions of the sequence, including nt 56–98, were determined only tentatively. Moreover, the consensus G100/G76 pairing is found in a closely related organism, *Proteus shigelloides*, sequenced more recently using more modern methods (Macdonnell & Colwell, 1985). Therefore, it is reasonable to exclude the G100/A76 pairing reported for *P. vulgaris* from consideration.

A99/U77

This *trans*-Hoogsteen pairing is strictly conserved in the database, and, as already noted, the geometry and reactivity are exactly as observed for U103/A73.

G98/A78

As already noted, this sheared purine–purine base pair is equivalent in geometry to A104/G72 and, as also found for the sheared A104/G72 pairing, the consensus G98/A78 covaries primarily with A/A (Table 2). As also found for A104/G72, a small number of sequences in which G is replaced by C or U are observed. The one G/G pairing found in the database is from the *P. vul-*

garis sequence discussed above. For the related *P. shigelloides*, the consensus G/A pairing is reported. Thus, the G/G pairing from *P. vulgaris* may also be removed from consideration. A single A98/U78 pairing is reported in the database, belonging to *Mycoplasma gallisepticum*, strain A5969. The A98/U78 pairing is not isosteric with G98/A78. However, this sequence was determined at the RNA level (Rogers et al., 1985) and more recent sequencing at the DNA level of a second variant of the same strain reports the isosteric A98/A78 for this pairing (Szymanski et al., 1998).

C97/G79

This is the sole Watson–Crick pair within loop E. All Watson–Crick pairs plus U/G (but no G/U) are observed at this position.

G96/U80

This pairing exhibits typical wobble geometry in which GN1 is H-bonded with UO2 and GO6 is H-bonded with UN3. G96 is stacked on G81 in a cross-strand manner similar to A104 stacking on A73 and A99 stacking on A78. This pairing is strictly conserved except for a small number of G/C substitutions. A correlation is observed with the next base pair (95/81). Whenever G/C occurs at 96/80, a Watson–Crick pairing also occurs at 95/81. The converse is not true, however.

U95/G81

This pairing shows a complex set of covariations, including Watson–Crick pairings, U/U, C/U, and C/C (not shown in Table 2). The distance of the hairpin loop from the pairing varies considerably and is correlated with the sequence variation observed. In the extreme case of the shortest sequences, this pairing is G95/C81 and actually comprises the closing Watson–Crick pair of the GNRA hairpin loop found in many 5S RNAs in loop B. A continuous variation in length is observed up to the full-length version comprising 5 bp between the 95/81 pairing and the hairpin loop, as exemplified by the *E. coli* sequence. The pyrimidine–pyrimidine pairings observed can be accommodated geometrically, but probably require bridging by water molecules to satisfy the H-bonding of the Watson–Crick faces, as observed in crystal structures of oligonucleotides having Y/Y mismatches (Holbrook et al., 1991; Cruse et al., 1994).

In summary, the conservative substitutions identified for each of the noncanonical pairings of loop E can be understood in terms of the 3D structure. Almost all can form pairings isosteric, or nearly isosteric, with the pairing observed in the crystal structure. Thus, in sheared A/G pairs, the G residue can be replaced by an A, C, or U, giving sheared A/A, A/C, and A/U pairs. The *trans*-Hoogsteen U/A pair almost never presents any conservative change, except for a handful of C/A pairs. The bifurcated G/U pair and the related G/G pair can be replaced by bifurcated A/A or A/C pairs. Finally, the open A/G pair exchanges with the alternative G/A open pair.

Concerted changes within the motifs of loop E

In a number of instances, concerted changes are observed in loop E. Some of them are confined to individual phylogenetic groups, whereas others are observed in more than one group. Some examples are shown in Figure 8.

Thermomicro roseum

An extra G in submotif 1, either between positions 71 and 72 or between 72 and 73, is accompanied by replacement of G102/U74 with a C/G base pair. The rest of loop E in this 5S RNA is identical to the consensus. This indicates some degree of functional independence between submotifs 1 and 2 of loop E.

Lactococcus lactis cremoris

In this 5S rRNA, changes are also localized to the upper half of loop E (Fig. 8). Submotif 2 appears intact, whereas submotif 1 appears to be shifted one pairing position away from the three-way junction. This is suggested by the occurrence of a U/G base pair where the A/G sheared is expected and the presence of an A/A pairing at position 101/75, where an open A/G or G/A is expected. Given that A/A can substitute isosterically for G102/U74, the three pairings A/G, U/A, and A/A should be considered a single unit. Thus, in this 5S rRNA, it is likely that the open A/G is deleted and submotifs 1 and 2 are in direct contact.

Anaerorhabdus furcosus

In this 5S, G102/U74 is replaced by C/U and A101/G75 by A/U (not shown). The substitution of A/U at this position is always accompanied by a change at the neighboring pair 102/74. For example, in *Mycoplasma genitalium* and *M. pneumoniae*, A101/U75 occurs with A102/A74.

Chromatium minutissimum

The entire upper half of loop E appears to be changed in this 5S rRNA as a result of one or more deletions (Fig. 8). It is not clear which bases are paired, much less what the structure is. Interestingly, the lower half of loop E is very similar to that of *E. coli*, to which *Chromatium* is related phylogenetically (Proteobacteria gamma).

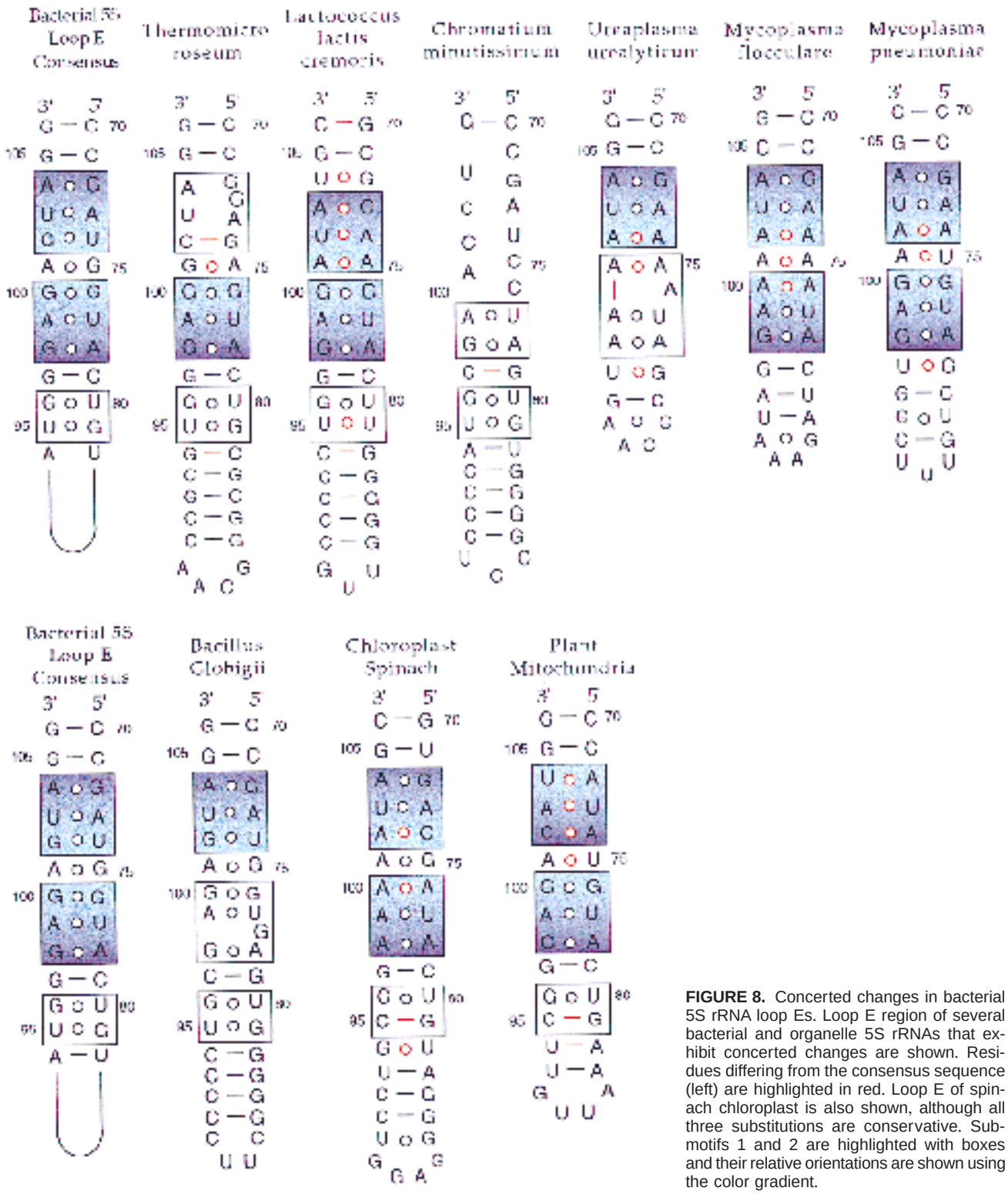

FIGURE 8. Concerted changes in bacterial 5S rRNA loop Es. Loop E region of several bacterial and organelle 5S rRNAs that exhibit concerted changes are shown. Residues differing from the consensus sequence (left) are highlighted in red. Loop E of spinach chloroplast is also shown, although all three substitutions are conservative. Submotifs 1 and 2 are highlighted with boxes and their relative orientations are shown using the color gradient.

Ureaplasma urealyticum

The deletion of a base 3′ to A99 produces an asymmetric loop E in this molecule. It is possible that A76 is bulged out, just as G76 is bulged in loop E's of eucaryal 5S rRNAs. A75 and A102 could then form a parallel symmetrical Hoogsteen–Hoogsteen pairing as in eucaryal loop E's. Thus, the loop E of this 5S has the potential to fold like a eucaryal 5S loop E. This example is interesting because it shows how closely related

bacterial and eucaryal loop E's actually are. Note that the upper half (submotif 1) of this loop E is unchanged (the observed A102/A74 pairing being isosteric with the bifurcated G102/U74 pair; Fig. 5).

Mycoplasma flocculare

This sequence differs from the preceding only by the presence of A100 (Fig. 8). It demonstrates explicitly the geometrical identity of the two symmetrical submotifs 1 and 2. Both the 102/74 and 100/76 pairings are A/A. Pairing 101/75 is also A/A, which, as discussed above, requires a small adjustment in H-bonding relative to the isosteric A/G and G/A pairings that normally occur at this position. This is perhaps related to changes in hydration at the neighboring positions. Another example having A101/A75 is found in *B. acidocaldarius*, in which G/G occurs at position 102/74 (recall that G/G is also isosteric with G102/U74; Fig. 5).

Mycoplasma pneumoniae

In this sequence (and also in the closely related *M. genitalium*), the normal G/G pairing occurs at 100/76, whereas A/U occurs at 101/75 and A/A at 102/74 (Fig. 8). A102/A74 is characteristic of mycoplasma 5S rRNAs. However, other mycoplasmas have the normal A/G or G/A pairings at position 101/75.

Bacillus globigii

This provides an example of an insertion between U77 and A78 (Fig. 8, lower panel), creating again the potential for a eucaryal 5S rRNA loop E type structure. The inserted G could pair with A99 to create a tandem 5′-GA-3′/5′-GA-3′ pair (geometrically compatible with 5′-UA-3′/5′-GA-3′) and bulging U77 at a position equivalent to G76 in eucaryal loop E.

Plant chloroplasts

The loop E of a typical chloroplast (from spinach) is shown in Figure 8 (lower panel). The characteristic features include substitution of A/C for G/U at position 102/74, A/A for G/G at 100/76, and sheared A/A for G/A at 98/78. All of these are conservative substitutions observed in other sequences. Note that chloroplast and mitochondrial sequences were excluded from the analysis, which led to the identification of these conservative substitutions. In some chloroplast sequences, the non-isosteric A/A pairing occurs at 101/75, as noted above for mycobacteria.

Plant mitochondria

In plant mitochondria (also excluded from the phylogenetic analysis above), the four noncanonical base pairs closest to the three-way junction are replaced as shown in the lower panel of Figure 8. The rest of loop E is identical to the consensus, with the exception of pairing 95/81, which is Watson–Crick. We were able to model A72/U104, which replaces the sheared G72/A104, as a Watson–Crick pair, while keeping a *trans*-Hoogsteen conformation for the adjacent U73/A103. C102/A74 can have the same conformation as a sheared G/A pairing. This analysis indicates that it is possible that submotif 1 is rotated 180° in these sequences, as shown schematically by the color gradient in Figure 8.

Chemical probing of spinach chloroplast 5S: A test of the universality of the loop E structure

The 5S rRNA of spinach chloroplast has also been the subject of careful chemical probing and molecular modeling. We review those data to test consistency with conservation of the 3D structure in this group of 5S rRNAs. The spinach chloroplast sequence is representative of other chloroplast 5S rRNAs, which are characterized by conservative substitutions at three positions. The chemical probing data on the loop E region of the 5S rRNA from spinach chloroplasts (Romby et al., 1988; Westhof et al., 1989) are summarized in Figure 9. Note that the chloroplast 5S was probed at 37 °C, whereas the *E. coli* data presented in Figure 2 were obtained at 20 °C. This helps to explain the generally greater reactivity of the loop E residues of chloroplast 5S to chemical probes compared with that of *E. coli* 5S rRNA. Many of the positions that are only reactive under denaturing conditions in the *E. coli* 5S (blue in Figs. 2 and 9) are reactive under semi-denaturing conditions in the chloroplast version of the molecule (green). Correspondingly, positions reactive under semi-denaturing conditions in *E. coli* 5S are reactive under native conditions in the chloroplast 5S. (For both data sets, native conditions comprised neutral buffers containing magnesium ions, whereas the ions were omitted from buffers for semi-denaturing conditions.) A further factor contributing to the greater reactivity of the chloroplast 5S is the larger proportion of AU and GU pairings in helix IV (the hairpin stem) and the substitution of three potentially less stable pairings within loop E itself: A/C for G102/U74, A/A for G100/G76, and A/A for G98/A78. Nonetheless, the same *relative* chemical reactivities are observed in the chloroplast 5S as for the *E. coli* molecule. For example, A101 is more reactive than G75 at the N1 position in both chloroplast and *E. coli* 5S rRNAs. The same applies to A104 and G72. A78 and A104, both of which are involved in sheared purine pairings, and A73 and A99, both of which are involved in U/A *trans*-Hoogsteen pairings, are reactive under native conditions, in both chloroplast and *E. coli* 5S. The N7 of each of these adenosines is unreactive under native conditions. These data indicate that the 3D

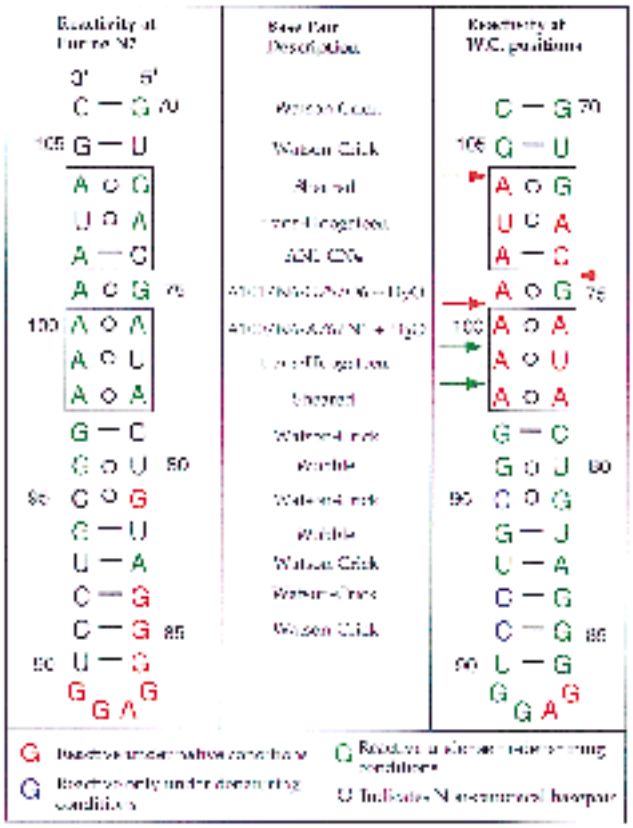

FIGURE 9. Chemical probing of spinach chloroplast 5S rRNA (Romby et al., 1988). Color coding is as in Figure 2. Green arrows indicate phosphates strongly protected and red arrows indicate phosphate somewhat protected compared to the average.

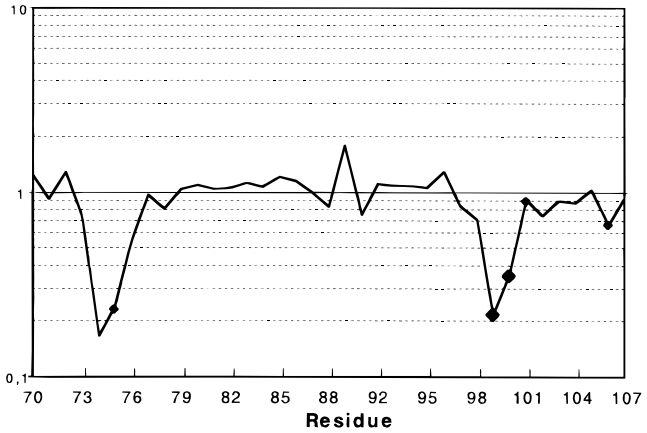

FIGURE 10. Plot of normalized accessibilities of phosphate groups (calculated as the average of the two nonbridging oxygen atoms) in loop E. Residues found to be more protected than the average in chemical probing experiments on chloroplast 5S rRNA (Romby et al., 1988) are marked with diamonds (large markers correspond to green arows in Fig. 9; small markers to red arrows).

structure of the chloroplast loop E is likely to be very similar to that of bacterial loop E.

The phosphates of G75, A99, A100, A101, and G105 were found to be protected from ethylnitrosourea relative to other phosphates of chloroplast 5S rRNA. A99 and A100 (green arrows in Fig. 9) were the most strongly protected. The accessibilities of the phosphates in the loop E crystal structure (relative to the average value of all phosphates in the structure) are shown in Figure 10, plotted as a function of position in the sequence. The correspondence of the chemical probing data to the calculated accessibilities is further evidence for the conservation of the 3D structure of loop E across bacterial and plastid phylogenetic groups. A 3D model of loop E of spinach chloroplast 5S rRNA, constructed by isosteric replacement of G102/U74 by A/C, G100/G76 by A/A, G98/A78 by A/A, and C97/G79 by G97/C79, is shown in stereo in Figure 11.

Loop E motif: Identification in other structured RNAs

In the analysis performed above, conservative substitutions at each of the noncanonical pairings within the loop E structure were identified in the database of bacterial 5S rRNA. These substitutions were found to be isosteric, or nearly so, to the corresponding base pairings observed in the structure, with only a handful of exceptions. This suggests the possibility of examining conservative substitutions within symmetric loops of other structured, functionally conserved RNA molecules with the goal of inferring 3D structure. In particular, it suggests the possibility of identifying the bacterial 5S loop E motif (or its submotifs) in other molecules. The experience with the closely related loop E motif of eucaryal 5S rRNA suggests that it is worth trying. The eucaryal loop E, which is identical to the sarcin–ricin loop motif in 23S rRNA, has been found in a large variety of RNA molecules, not only in the context of internal loops, but also within multi-helix junctions (N.B. Leontis & E. Westhof, 1998).

In fact, a likely candidate for a bacterial loop E submotif occurs in 16S rRNA at residues 581–583 and 758–760 in the *E. coli* 16S rRNA sequence. This is a symmetrical internal loop in bacteria and archae and most eucarya. In all phylogenetic groups, a Watson–Crick base pair flanks the motif on one side and a conserved GU on the other (Fig. 12). In bacterial sequences, the 583/758 pairing is 80.6% A/G (2892 sequences), 14.5% A/C, 3.2% A/A, and 1.06% A/U, all of which are isosteric substitutions observed for the sheared A/G pairings in 5S loop E, as discussed above. The neighboring 582/759 pairing is restricted to U/A (72.5%) and C/A (26.7%), the two nearly isosteric pairings identified in the 5S loop E for the *trans*-Hoogsteen U/A pair. Chemical probing indicates that the two conserved adenosines, A583 and A759, are reactive at their N1 (Watson–Crick) positions (Moazed et al., 1986), consistent with the proposed structure. The 581/760 pairing displays G/G (81.0%), G/U

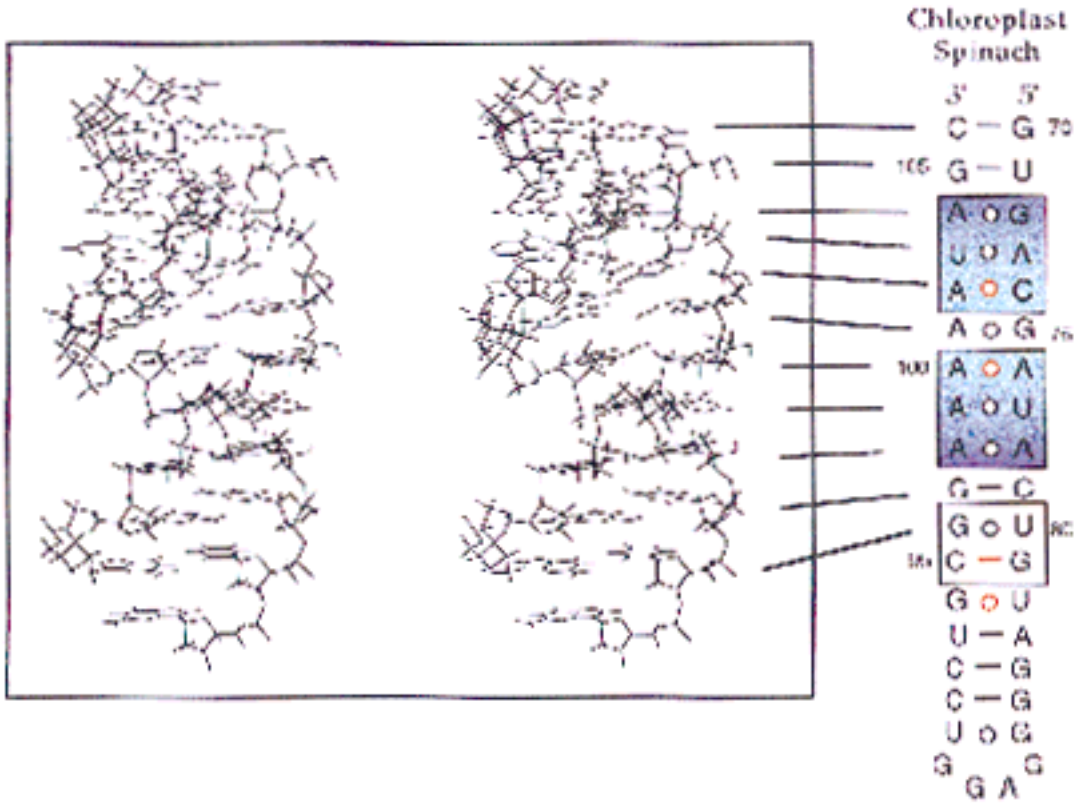

FIGURE 11. Three-dimensional model of the loop E of spinach chloroplast 5S rRNA generated from the 3D structure of *E. coli* by isosteric replacements of G102/U74 by A/C, G100/G76 by A/A, and G98/A78 by A/A.

(14.6%), G/A (3.0%), and C/A (0.8%) pairings. All these variations (except the rare C/A pairing) occur as conservative variations at the equivalent positions of loop E submotifs 1 or 2 and the two most abundant (G/G and G/U) are essentially isosteric, as shown in Figure 5. The most common bacterial versions of the 16S internal loop are shown in Figure 12. Boxes are drawn around those versions of the motif identical to submotifs 1 and 2 of the consensus 5S loop E.

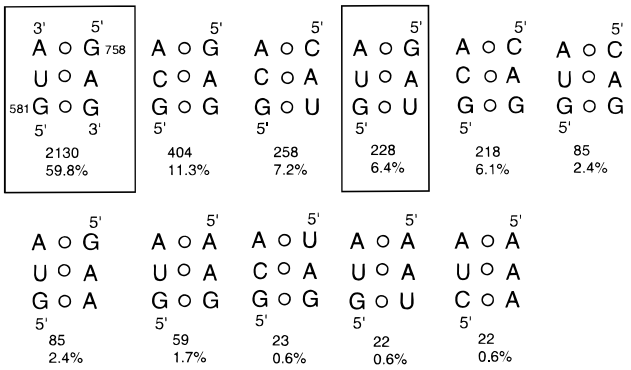

FIGURE 12. Most commonly occurring variations of an "internal loop" within bacterial 16S rRNA (residues 581–583 and 758–760 in the *E. coli* 16S rRNA sequence). The most abundant (59.8% of all bacterial sequences) corresponds exactly to submotif 2 of loop E (boxed). Submotif 1, although less abundant (6.4%) is also present (boxed).

In the *E. coli* 16S rRNA, the motif occurs with substitution of A/C for the sheared A/G and C/A for the *trans*-Hoogsteen U/A, a version of the motif that occurs in 6.1% of bacterial sequences. Physical evidence that this sequence likely shares the same 3D structure as the loop E motif comes from NMR spectroscopy (Lukavsky et al., 1997) and chemical probing studies (Lentzen et al., 1996) of an identical internal loop found in the conserved domain IV of the signal recognition particle (SRP) RNA. The NMR study shows that the adenosine bases of the A/C and C/A pair are stacked in the same cross-strand fashion found in submotifs 1 and 2 of loop E. The chemical probing study shows that these same adenosines are reactive at their N1 positions. Moreover, it was found that the N1 of the guanosine corresponding to G100 in 5S rRNA is reactive, whereas that of its partner, which corresponds to G76 in loop E, is not, exactly as observed in loop E, providing support for a bifurcated G/G pairing in the SRP internal loop.

CONCLUSIONS

1. The original conclusions regarding the structure of loop E based on chemical probing more than 10 years ago

were essentially correct. Loop E is highly structured, with extensive base pairing within the loop, and magnesium ions play a key role in that structuring (Romby et al., 1988; Westhof et al., 1989; Brunel et al., 1991). However, it is now apparent that the key role played by water and ions in protecting positions not involved in direct RNA–RNA H-bonds and particularly the direct involvement of water molecules in completing the H-bonding in some pairings render the process of deriving a given base pairing scheme from chemical probing data extremely difficult, except for *trans*-Hoogsteen A/U and sheared A/G pairs, which possess clear signatures in chemical probing.

Thus, the cautious view reached after the present analysis is rather different from that attained after the extensive comparisons between chemical and enzymatic probing and crystallographic structures of tRNAs made several years ago (Holbrook & Kim, 1983; Romby et al., 1985). In tRNAs, also, the phosphate protections were rather well reproduced by the calculations based on the structures (Romby et al., 1985). Interestingly, the most frequent non-Watson–Crick pair in tRNAs is the *trans*-Hoogsteen A/U pair, for which we see a clear-cut chemical probing signature.

2. To evaluate the range of possible isosteric substitutions existing in a conserved molecule such as 5S rRNA, it is necessary to separate conservative sequence changes at the single base pair level from concerted changes involving several base pairs. The large majority of conservative substitutions appear to be capable of forming isosteric replacements requiring little or no geometrical readjustment relative to the crystal structure. In most cases, identical hydration patterns can also be proposed for these substitutions, emphasizing the integral part of water molecules in RNA structures.

3. The G100/G76 and G102/U74 bifurcated pairings are essentially identical in geometry. The overlapping covariations observed in the database, particularly the occurrence of A/A at both positions, provides evidence that these may substitute for each other in other occurrences of the loop E motif.

4. Loop E comprises semi-independent submotifs. This is supported by the observation that in most loop E's having concerted changes, these are localized to one or the other of submotifs 1 and 2. The *U. urealyticum* and the *Bacillus globigii* sequences demonstrate that deletion of one base may transform a submotif of a bacterial loop E into the related eucaryal loop E motif.

5. Analysis in the light of phylogenetic data of the structures of noncanonical pairings of "internal loops" (such as loop E) allows one to generate a dictionary of isosteric substitutions for noncanonical pairings. These in turn can be applied to postulate 3D structures for other internal loops that show overlapping patterns of sequence variation. This analysis supports the suggestion that at least one internal loop in 16S and another in SRP RNA (4.5 S RNA in bacteria) share a common structure with submotifs 1 and 2 of loop E of bacterial 5S rRNA.

MATERIALS AND METHODS

The 5S rRNA crystal structures (accession numbers URL064, URL066, and URL069) were obtained from the Nucleic Acids Database (http://ndb.rutgers.edu/NDB/ndb.html). Accessibility calculations were performed using the Access program based on the algorithm of Richmond (1984) and sequence covariation analysis using the COSEQ program.

ACKNOWLEDGMENTS

We are grateful to Christian Massire for help with the use of the COSEQ program. We thank Bernard and Chantal Ehresmann, Pascale Romby, and Christine Brunel for numerous, long-standing, and insightful discussions and contributions to our understanding of the relationship between chemical probing and structure in 5S rRNAs. N.L. was a Fogerty Senior International Fellow, supported by NIH grants 1-F06-TW02251-01 and 1R15-GM/OD55898-01, ACS-PRF grant #31427-B4, and by a "Poste Rouge" from the CNRS. E.W. acknowledges support from the Institut Universitaire de France.

Received April 6, 1998; returned for revision May 8, 1998; revised manuscript received June 15, 1998

REFERENCES

Auffinger P, Louise-May S, Westhof E. 1996. Molecular dynamics simulation of the anticodon hairpin of tRNAasp: Structuring effects of C-HO hydrogen bonds and of long-range hydration forces. *J Am Chem Soc 118*:1181–1189.

Auffinger P, Westhof E. 1997. RNA hydration: Three nanoseconds of multiple molecular dynamics simulations of the solvated tRNA Asp anticodon hairpin. *J Mol Biol 269*:326–341.

Auffinger P, Westhof E. 1998. Effects of pseudouridylation on tRNA hydration and dynamics: A theoretical approach. In: Grosjean H, Benne R, eds. *Modification and editing of RNA*. Washington, DC: ASM Press. pp 103–112.

Brunel C, Romby P, Westhof E, Ehresmann C, Ehresmann B. 1991. Three-dimensional model of *E. coli* ribosomal 5S RNA as deduced from structure probing in solution and computer modeling. *J Mol Biol 221*:293–308.

Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA. 1996. Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science 273*:1678–1685.

Chiu DK, Kolodziedjczak T. 1991. Inferring consensus structure from nucleic acid sequence. *CABIOS 7*:347–352.

Correll CC, Freeborn B, Moore PB, Steitz TA. 1997. Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell 91*:705–712.

Cruse WBT, Saludjian P, Biala E, Strazewski P, Prangé T, Kennard O. 1994. Structure of a mispaired RNA double helix at 1.6-Å resolution and implications for the prediction of RNA secondary structure. *Proc Natl Acad Sci USA 91*:4160–4164.

Dallas A, Moore PB. 1997. The loop E loop D region of *Escherichia coli* 5S RRNA: The solution structure reveals an unusual loop that may be important for binding ribosomal proteins. *Structure 5*(12):1639–1653.

Davis DR. 1998. Biophysical and conformational properties of modified nucleosides in RNA (NMR studies). In: Grosjean H, Benne R, eds. *Modification and editing of RNA*. Washington DC: ASM Press. pp 85–102.

Dennisov VP, Carlström G, Venu K, Halle B. 1997. Kinetics of DNA hydration. *J Mol Biol 268*:118–136.

Fischel JL, Ebel JP. 1975. Sequence studies on the 5S RNA of *Proteus vulgaris*: Comparison with the 5S RNA of *Escherichia coli*. *Biochimie 57*:899–904.

Fox GE, Woese CR. 1975. 5S RNA secondary structure. *Nature 256*:505–507.

Holbrook SH, Kim SH. 1983. Correlation between chemical modification and surface accessibility in yeast phenylalanine transfer RNA. *Biopolymers 22*:1145–1166.

Holbrook SR, Cheong C, Tinoco I, Kim SH. 1991. Crystal structure of an RNA double-helix incorporating a track of non-Watson–Crick base pairs. *Nature 353*(10):579–581.

Kim SH, Quigley GJ, Suddath FL, McPherson A, Sneden D, Kim JJ, Weinzierl J, Rich A. 1973. Three-dimensional structure of yeast tRNA(Phe): Folding of the polynucleotide chain. *Science 179*:285–288.

Lentzen G, Moine H, Ehresmann C, Ehresmann B, Wintermeyer W. 1996. Structure of 4.5 S RNA in the signal recognition particle of *E. coli* as studied by enzymatic and chemical probing. *RNA 2*:244–253.

Leontis NB, Ghosh P, Moore PB. 1986. Effect of magnesium ion on the structure of the 5S RNA form *Escherichia coli* and imino proton magnetic resonance study of the helix I, IV, and V regions of the molecule. *Biochemistry 25*:7386–7392.

Leontis NB, Westhof E. 1998. A common motif organizes the structure of multi-helix loops in 16S and 23S ribosomal RNAs. *J Mol Biol*. In press.

Lukavsky P, Billeci TM, James TL, Schmitz U. 1997. NMR structure determination of a 28-nucleotide signal recognition particle RNA with complete relaxation matrix methods using corrected NOE intensities. In: Leontis NB, Santa Lucia J, eds. *Molecular modeling of nucleic acids*. Washington, DC: American Chemical Society. pp 122–149.

Macdonnell MT, Colwell RR. 1985. Phylogeny of the vibrionaceae. *System Appl Microbiol 6*:171–182.

Moazed D, Stern S, Noller HF. 1986. Rapid chemical probing of conformation in 16S ribosomal RNA and 30S ribosomal subunits using primer extension. *J Mol Biol 187*:399–416.

Noller HF. 1998. Ribosomal RNA. In: Simons RW, Grunberg-Manago M, eds. *RNA structure and function*. Plainview, New York: Cold Spring Harbor Laboratory Press.

Richmond TJ. 1984. Solvent accessible surface area and excluded volume in proteins. *J Mol Biol 178*:63–89.

Rogers MJ, Simmons J, Walker RT, Weisburg WG, Woese CR, Tanner RS, Robinson IM, Stahl DH, Olsen G, Leach RH, Maniloff J. 1985. Construction of the mycoplasma evolutionary tree from 5S rRNA sequence data. *Proc Natl Acad Sci USA 82*:1160–1164.

Romby P, Baudin F, Brunel C, Leal de Stevenson I, Westhof E, Romaniu PJ, Ehresmann C, Ehresmann B. 1990. Ribosomal 5S RNA from *Xenopus laevis* oocytes: Conformation and interaction with transcription factor IIIA. *Biochimie 72*:437–452.

Romby P, Moras D, Bergdoll M, Dumas P, Vlassov VV, Westhof E, Ebel JP, Giegé R. 1985. Yeast tRNA(ASP) tertiary structure in solution and areas of interaction of the tRNA with aspartyl-tRNA synthetase. *J Mol Biol 184*:455–471.

Romby P, Westhof E, Toukifimpa R, Mache R, Ebel JP, Ehresmann C, Ehresmann B. 1988. Higher order structure of chloroplastic 5S ribosomal RNA from spinach. *Biochemistry 27*:4721–4730.

Saenger W. 1984. tRNA—A treasury of stereochemical information. In: *Principles of nucleic acid structure*. New York: Springer Verlag. pp 331–349.

Sussman JL, Holbrook SR, Warrant RW, Church GM, Kim SH. 1978. Crystal structure of yeast phenylalanine tRNA. I. Crystallographic refinement. *J Mol Biol 123*:607–630.

Szewczak AA, Moore PB. 1995. The sarcin/ricin loop, a modular RNA. *J Mol Biol 247*:81–98.

Szymanski M, Specht T, Barciszewska MZ, Barciszewski J, Erdmann VA. 1998. 5S rRNA data bank. *Nucleic Acids Res 26*:156–159.

Walczak R, Westhof E, Carbon P, Krol A. 1996. A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs. *RNA 2*:354–366.

Westhof E. 1988. Water: An intergral part of nucleic acid structure. *Annu Rev Biophys Biophys Chem 17*:125–144.

Westhof E. 1993. Modelling the three-dimensional structure of ribonucleic acids. *J Mol Struct 286*:203–210.

Westhof E, Romby P, Romaniuk PJ, Ebel JP, Ehresmann C, Ehresmann B. 1989. Computer modelling from solution data of spinach chloroplast and of *Xenopus laevis* somatic and oocyte 5S rRNAs. *J Mol Biol 207*:417–431.

Wimberly B, Varani G, Tinoco I Jr. 1993. The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochemistry 32*(4):1078–1087.